Essential Components of Writing Assessments

Carolee Gunn, Ph.D.

Presentation at the Montana OPI

Assessment Conference 4/28/08

Essential Elements

- Clear statement of assessment's purpose
- Strong validity evidence
- Reasonable levels of score reliability
- Alignment with content standards
- Adequate technology
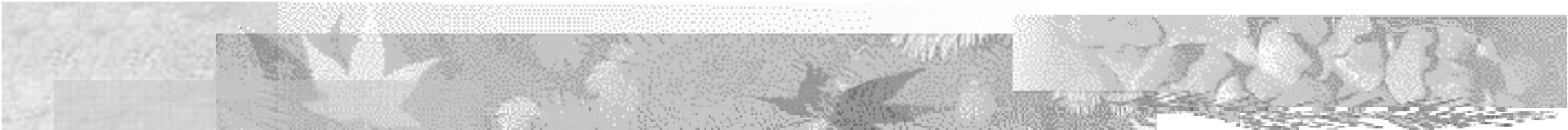- Sufficient resources

Purpose of the writing assessment: rationale and objectives

To determine:

- The individual student's skills & needs
- The effectiveness of instruction
- Readiness for post-secondary education, training, and employment

And/Or to meet federal & state accountability

What is the best approach to assessment given the purpose of the test?
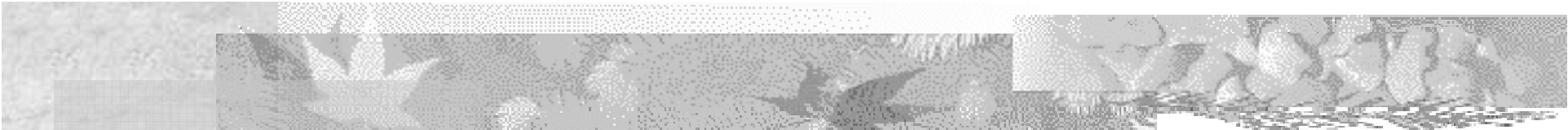
Formative assessment

- Valuable for guiding instruction and providing feedback to students

- Frequent assessments aligned with current unit of instruction

- Repeated cycles of scoring and reporting

- Generally not aggregated beyond the class

What is the best approach to assessment given the purpose of the test?
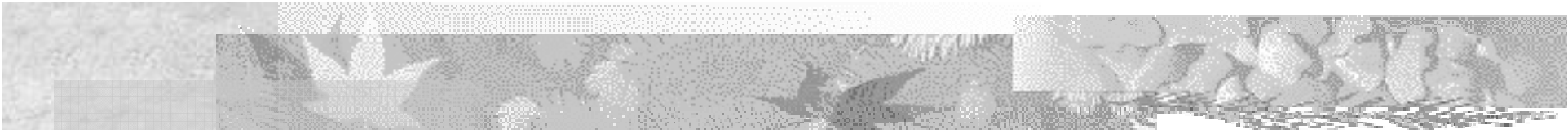
Interim assessment

- Medium-scale, medium-cycle assessments
- Designed to inform decisions at the classroom, school and district levels.
- Results may be meaningfully aggregated and reported at a broader level than classroom
- Less instructionally relevant than formative

What is the best approach to assessment given the purpose of the test?
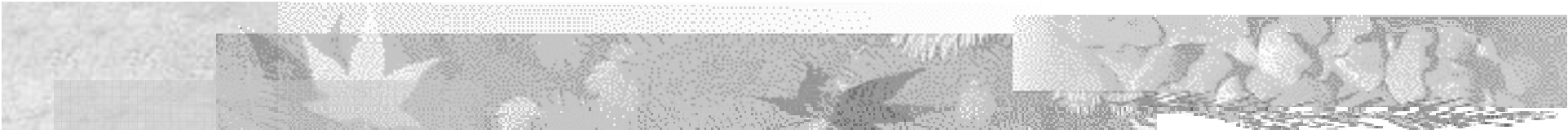
Summative assessment

- Infrequent assessment
- Less time consuming than formative
- Limited feedback to students & teachers
- Scores can be aggregated to the school, district and state levels
- Aligned to state content standards

How broadly should the test be administered?

National: The National Assessment of Educational Progress (NAEP)
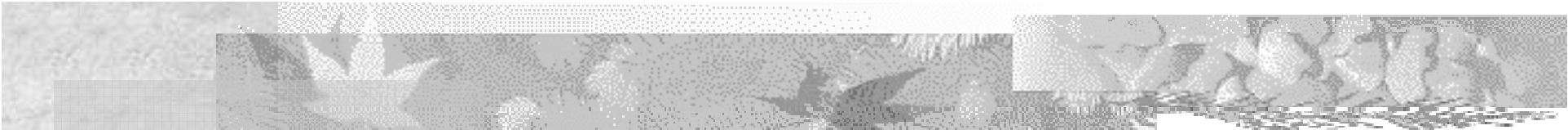
- Administered across the country to a representative sample of students in grades four, eight and twelve.

- The writing assessment is administered every four years.

- In 2007, only grades 8 & 12 participated in the writing assessment in Montana.

How broadly should the test be administered?

Statewide

- The majority of states administer statewide assessments except those with a preference for local tests, such as Montana, Colorado, Iowa and Nebraska.

- Some states include writing tests to determine their statewide adequate yearly progress in reading.

How broadly should the test be administered?

Local Assessments

- Designed to meet local needs and to stay within local jurisdiction

- Are they more likely to be formative than summative?

- Are the teachers more invested in local assessments?

What grade levels should be targeted?

- A NAEP study conducted in 2006 revealed that for the majority of states, the lowest grades tested were grades 4 & 5. Only four states tested students in grade 3.

- At the middle school level: grades 7 &8

- At the high school level: grades 10 & 11

Defining the construct of writing

- The construct definition should reflect the state's academic content standards at the target grade levels.
- The construct definition varies state-by-state in terms of depth and breadth, i.e., do the students evaluate written text or create their own text?
- Does the construct set the writing medium?

Genres of Writing

The traditional genres characterize the *types* of writing.

- **Narrative**: fictional stories, personal essays, diaries, etc.

- **Informative**: explanation and analysis

- **Descriptive:** painting a picture in words

- **Persuasive**: argument and debate

Purposes for writing

Good writing is constructed with a goal or objective in mind.

Effective communication must take into account specific audiences and purpose.

The ability to make appropriate choices in relation to purpose and audience is essential in school and in the workplace.

Source: NAEP 2011 Writing Framework

Purposes for Writing

Innovative approach proposed in the NAEP writing framework for 2011

- <u>To persuade</u> in order to change the reader's opinion or to affect the reader's actions
- <u>To explain</u> in order to expand the reader's understanding
- <u>To convey</u> experience, real or imagined

Identifying the Audience

- Is the default audience always the teacher?!
- Variance among elementary, high school, and middle school students' ability to take another's perspective ranges from "I don't know" to "I don't care" to "Huh?"
- Be mindful of cultural differences in the appropriateness of youth trying to persuade authorities. Is that disrespectful?

Form of the Student's Response

Response formats:essay, letter, story, editorial, report, or even an e-mail message

- Should the response mode be specified or should the students choose the form that suits their purpose?

- Does form specification make a difference in time-limited assessments?

- Should the form be specified for some students who have cognitive disabilities?

Topic of the writing prompt

- The challenge: to select topics that are familiar across all student subgroups, and spark an interest in writing.

- Are the topics age-appropriate and not offensive to any students?

- Is there a "topic effect" that strongly influences the students' responses?

Administration of the Assessment

Time limits for the student to respond

Shortest time: 25-30 minutes per prompt

Advantage: The students can respond to two prompts (NAEP model).

Challenge: Students must think fast and write fast, and have little time to revise.

Outcome: A draft rather than polished writing

Administration of the Assessment

Time limits for the student to respond

Longest time: one week

Advantage: the student's response is a polished product rather than a draft.

Challenge: The lack of standardization increases the score variance/error.

Question: How much assistance do the students receive?

Computer-Based Writing

Shouldn't the assessment reflect the way students and adults write now in the early 21st Century?

Examine the academic achievement standards for writing and word-processing skills. Are there overlapping skills?

Trends in computer-based writing

- NAEP writing assessment questionnaire in 2002 revealed that 97% of grade 8 teachers claimed their students used computers to compose written text.*

- In 2007, nineteen states were using, or developing direct writing assessments using computers (ACT Research Brief).

Trends in computer-based writing

- A paper & pencil assessment could create issues of bias for students who commonly use computers to write.*
- If a word processor is used for computer-based assessments, what tools should be enabled? Spelling and grammar checking?
- What is the relationship between the test medium and the construct definition?

Criteria for Evaluating Responses

- Incorporation of the six-traits components in scoring rubrics, based on the NWREL model: ideas, organization, voice, word choice, sentence fluency, and conventions.

- Holistic vs analytic scoring models/rubrics

Inter-correlation of the six writing traits

What does research reveal about the independence of the traits?

- The inter-correlation among the traits is very high when the data are aggregated to the school level.
- Only the correlations with the "conventions" component is significantly lower.

Analytic vs Holistic Scoring

- Since the different trait scores correlate highly, the holistic scoring approach may be more efficient and economical.

- Analytic scoring's value is maximized in a classroom-level formative assessment.

- Still, the criteria for scoring writing samples should include an evaluation of the components of this multidimensional skill.

NAEP Criteria for Evaluating Students' Writing

- Development of ideas is effective in relation to the writer's purpose and audience
- Organization is logical in relation to the writer's purpose and audience
- Language facility and conventions support clarity of expression and the effectiveness of the writing

Score Reliability

Approaches to maximizing score reliability

- Select highly-qualified individuals
- Generate a well-written scoring rubric
- Establish extensive scorer training
- *Backreading* by scoring supervisors
- Monitor scorer drift
- Provide retraining, as needed

Inter-Rater Agreement

Reasonable standards for perfect agreement:

- 70% agreement on a <u>holistic</u> 4-point scale

- 65% agreement on a <u>holistic</u> 6-point scale

- 60% agreement on an <u>analytic</u> 6x6 scale (six traits across a 6-point scale)

Automated scoring of writing

- Has artificial intelligence reached a level of sophistication that makes it a viable option for low-stakes testing?

- New terminology: automated essay scoring (AES) and latent semantic analysis

- AES enables fast and accurate scoring of formative assessments of writing skills.

Automated scoring of writing

Current AES *engines* can:

Determine if the text structure is close to standard English; identify grammatical errors and misspellings; evaluate the vocabulary sophistication; judge the sentence-to-sentence flow; and evaluate the overall coherence of the text.

Establishing Achievement Standards

- Has the state/district adopted challenging achievement standards in writing?
- Were the cut scores established using a recognized professional process?
- Are the performance-level descriptors detailed enough to inform parents and students of the expected achievement associated with each level?

Score Reporting

- Does the state/district provide individual interpretive reports that indicate relative strengths and weaknesses?

- Do the individual reports include clear information for the student and his/her parents?

- Are the reports accompanied by interpretive guides?

Slides posted by the OPI

- These presentation slides will be posted at
www.opi.mt.gov/AssessConf/index.html